User's Guide

Performance and Sizing Data

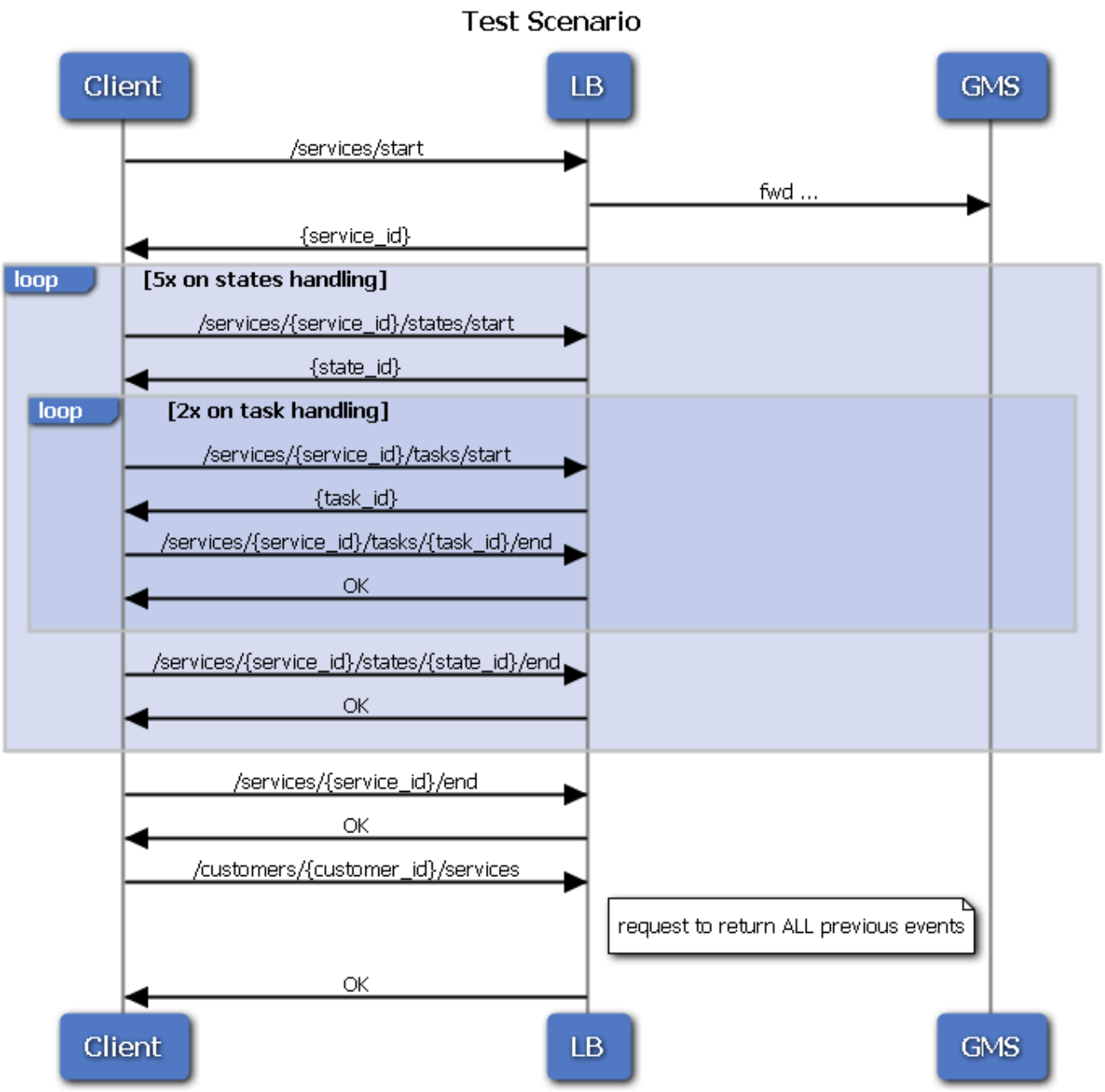
4/30/2025

Performance and Sizing Data

This page provides an example of sizing and performance for Context Services in a GMS deployment.

Contents

Sizing Scenario

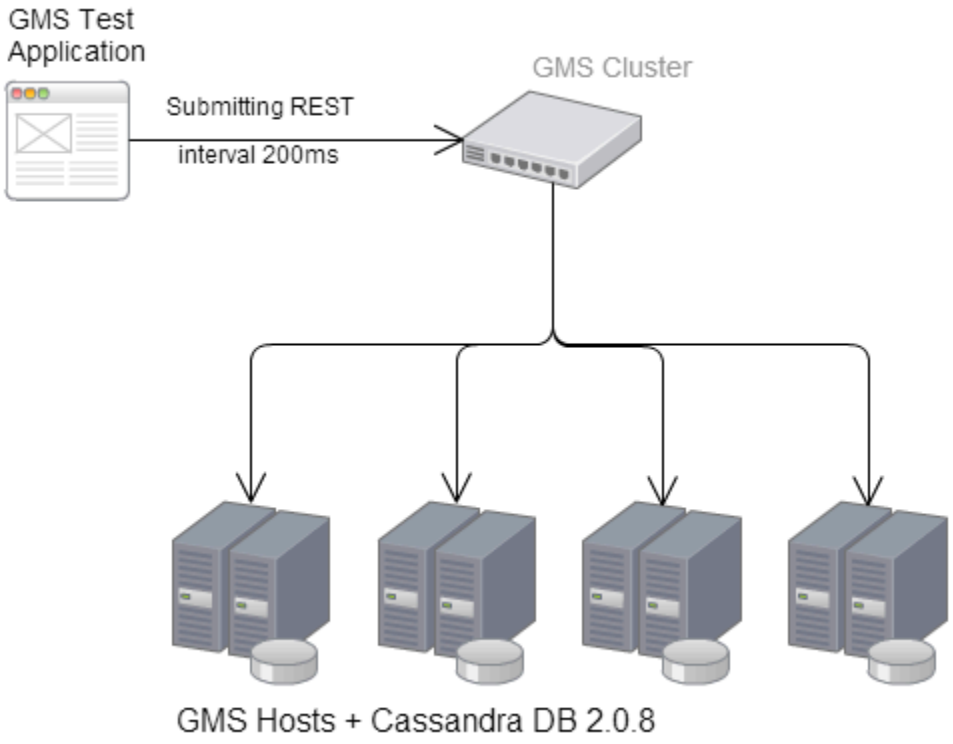


A Load Balancer (LB) receives the REST queries and dispatch them to GMS applications.

Our test scenario create a conversation implemented with one service including 5 states and tasks, which are equivalent to **33 REST queries** to handle on GMS side.

- 1 to 4 extension attributes are submitted in start events for services, states, and tasks.
- All the event attributes are set in Start and complete events using random values for integer a strings.

Sizing Settings



Number of hosts in use	4 (one per GMS cluster node)
Host Specifications	• Windows Server 2008 R2 Standard SP 1 • Intel Xeon X3440 @ 2.53 GHz, quad core • 8 GB RAM • Low Price disks
Components	• GMS 8.5.003.00+ • Cassandra 2.0.8 external
GMS Configuration	• Constant average CPU of 50% • Cassandra 2.0.8 (database of 2.5GB dat each node) • Replication Factor = 2

Sizing Results

Captured measure	Value	Comments
CPU	~20% to 60%	Each node was around an average of 30% with peaks to 60% from time to time.
Database size	68 GB / node (after 120h) = 272 GB	272 GB corresponds to 16.5 M conversations or 264 M service/ state/task events, for a ratio of 0.54KB per conversation (1 service, 5 states, 10 tasks).
I/O Disk	~3 MB /s (Cassandra measure)	Disk I/Os are fluctuating a lot (between 0 MB/s and 20 MB/s)
I/O Network	~3 MB /s	Between 20Mbps and 60Mbps.
Memory	2.4GB	For each GMS processes. Less than 2 GB for Cassandra processes.
Size of conversation (1 service, 5 states, 10 tasks)	~10kB in JSON	This matches a conversation (1 service, 5 states, 10 tasks), for 33 REST queries events, including extension data returned with queries by CustomerId.

Captured measure	Value	Comments
Throughput	~1000 queries/s (~30,5 scenario/s)	Constant through 24 hours of testing